0. Slack

Display name: JI YIHONG

1. Linear Algebra Review

1.1

$$\mathcal{L}(\vec{x},\lambda) = (\vec{x}-\vec{y})^T(\vec{x}-\vec{y}) + \lambda(\vec{\theta}^T\vec{x}+\theta_0)$$

$$= \sum_{i=1}^{d}(x_i-y_i)^2 + \lambda\left(\sum_{i=1}^{d}\theta_i x_i + \theta_0\right)$$

$$\frac{\partial L(\vec{x},\lambda)}{\partial x_i} = 2(x_i-y_i) + \lambda\theta_i \quad , \text{ where } i=1,2,\ldots,d$$

$$\frac{\partial L(\vec{x},\lambda)}{\partial \lambda} = \sum_{i=1}^{d}\theta_i x_i + \theta_0$$

or compactly

$$\frac{\partial \mathcal{L}}{\partial x_i} = i^{th} \text{ component of } 2(\vec{x}-\vec{y}) + \lambda\vec{\theta} \ , \ i=1,2,\ldots,d$$

$$\frac{\partial L}{\partial \lambda} = \vec{\theta}^T\vec{x} + \theta_0$$

1.2

By setting $\nabla_\theta \mathcal{L} = \vec{0}$, $\nabla_{\vec{x}} \mathcal{L} = \vec{0}$

$$2(\tilde{\vec{x}}-\vec{y}) + \lambda\vec{\theta} = 0 \quad — \ 1$$

$$\vec{\theta}^T\tilde{\vec{x}} + \theta_0 = 0 \quad — \ 2$$

Multiply both sides of ① by $\vec{\theta}^T$

$$2\vec{\theta}^T(\tilde{\vec{x}} - \vec{y}) + \lambda\vec{\theta}^T\vec{\theta} = 0$$

$$\Rightarrow \lambda = 2\,\frac{-\vec{\theta}^T\tilde{\vec{x}} + \vec{\theta}\vec{y}}{\vec{\theta}^T\vec{\theta}}$$

$$= 2\,\frac{\theta_0 + \vec{\theta}^T\vec{y}}{\vec{\theta}^T\vec{\theta}}$$ $(\vec{\theta}^T\vec{x} + \theta_0 = 0 \Rightarrow \vec{\theta}^T\vec{x} = -\theta_0)$

Replace λ in ①

$$\cancel{2}(\tilde{\vec{x}} - \vec{y}) + \cancel{2}\frac{\theta_0 + \vec{\theta}^T\vec{y}}{\vec{\theta}^T\vec{\theta}}\vec{\theta} = 0$$

Making $\tilde{\vec{x}}$ the subject

$$\tilde{\vec{x}} = \vec{y} - \frac{\theta_0 + \vec{\theta}^T\vec{y}}{\vec{\theta}^T\vec{\theta}}\vec{\theta}$$

<u>1.3</u>

Distance between $\vec{y}$ and H

$$= \|\vec{y} - \tilde{\vec{x}}\|$$

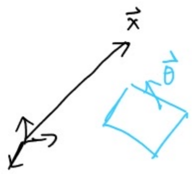
$$= \left\|\frac{\theta_0 + \vec{\theta}^T\vec{y}}{\vec{\theta}^T\vec{\theta}}\vec{\theta}\right\|$$

$$= |\theta_0 + \vec{\theta}^T\vec{y}| \cdot \frac{\|\vec{\theta}\|}{\vec{\theta}^T\vec{\theta}}$$

$$= \frac{|\theta_0 + \vec{\theta}^T\vec{y}|}{\|\vec{\theta}\|}$$ where $\|\vec{\theta}\| = \sqrt{\vec{\theta}^T\vec{\theta}}$

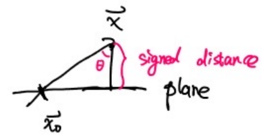
Alternative method using geometric method instead of Lagrange multiplier

Perpendicular distance to plane



$\vec{\theta}\cdot\vec{x}+\theta_0=0$

$\Rightarrow \theta_1 x_1+\theta_2 x_2+\cdots+\theta_n x_n+\theta_0=0$



$\vec{x_0}=c\vec{x}$,

$\vec{\theta}\cdot\vec{x_0}+\theta_0=0$

$\Rightarrow \vec{\theta}\cdot c\vec{x}+\theta_0=0$

$\Rightarrow c=\dfrac{-\theta_0}{\vec{\theta}\cdot\vec{x}}$

Hence, $\vec{x_0}=-\dfrac{\theta_0}{\vec{\theta}\cdot\vec{x}}\vec{x}$.

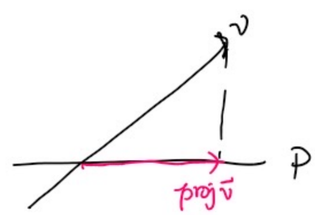
signed distance

$$=\|\vec{x}-\vec{x_0}\|\cos\theta$$

$$=\frac{\vec{\theta}\cdot(\vec{x}-\vec{x_0})}{\|\vec{\theta}\|}$$

$$=\frac{\vec{\theta}\cdot\left(\frac{\vec{\theta}\cdot\vec{x}+\theta_0}{\vec{\theta}\cdot\vec{x}}\right)\cdot\vec{x}}{\|\vec{\theta}\|}=\frac{\vec{\theta}\cdot\vec{x}+\theta_0}{\|\vec{\theta}\|}$$

Orthogonal Projection onto plane



$$\text{proj}_{\vec{v}}=\vec{v}-\text{perp}_{\vec{v}}$$

$$=\vec{v}-\frac{\vec{\theta}}{\|\vec{\theta}\|}\cdot\frac{\vec{\theta}\cdot\vec{v}+\theta_0}{\|\vec{\theta}\|}$$

2. Probability Review

2.1

$$\mathbb{P}(Z=z) = \mathbb{1}\{z \geq 0\} \cdot \sum_{x+y=z} \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)$$

$$= \underbrace{\mathbb{1}\{z \geq 0\}}_{\text{Indicator function}} \cdot \sum_{x=0}^{z} \mathbb{P}(X=x) \cdot \mathbb{P}(Y=z-x)$$

2.2

$$\mathbb{P}(Z=z) = \sum_{x=0}^{z} \frac{\alpha^{x} e^{-\alpha}}{x!} \cdot \frac{\beta^{z-x} e^{-\beta}}{(z-x)!}$$

$$= e^{-(\alpha+\beta)} \cdot \sum_{x=0}^{z} \frac{\alpha^{x} \beta^{z-x}}{x!\,(z-x)!}$$

$$= \frac{e^{-(\alpha+\beta)}}{z!} \cdot \sum_{x=0}^{z} \underbrace{\frac{z!}{x!\,(z-x)!}}_{\binom{z}{x}} \cdot \alpha^{x} \cdot \beta^{(z-x)}$$

$$= \frac{e^{-(\alpha+\beta)}}{z!} \cdot (\alpha+\beta)^{z} \quad \text{(binomial expansion)}$$

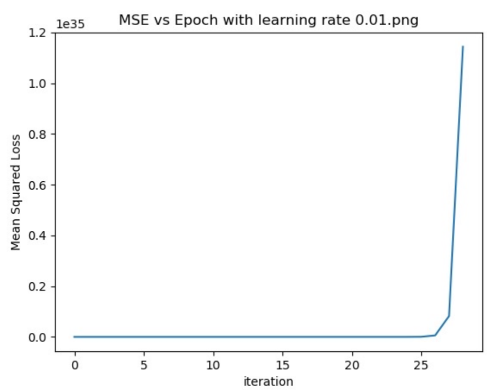
$$= \frac{(\alpha+\beta)^{z} e^{-(\alpha+\beta)}}{z!} \qquad \text{for } z \geq 0.$$

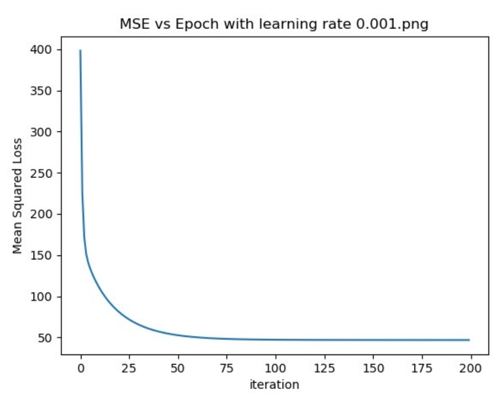
Hence, Z is also Poisson with rate $\varphi = \alpha + \beta$

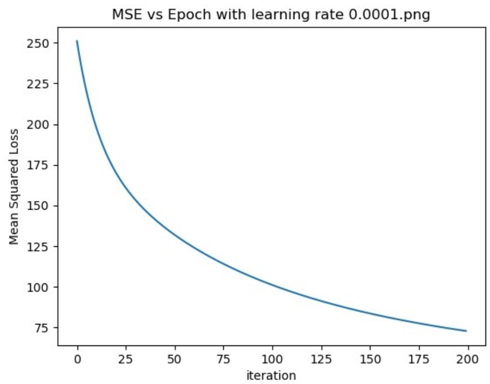
3. Linear Regression

$$\hat{y} = \underbrace{\begin{bmatrix} -\text{train } 1- \\ -\text{train } 2- \\ \vdots \\ -\text{train } 506- \end{bmatrix}}_{506 \times 3} \underbrace{\begin{bmatrix} W_{RM} \\ W_{RAD} \\ W_{CRIM} \end{bmatrix}}_{3 \times 1} + \underbrace{\begin{bmatrix} b \\ b \\ \vdots \\ b \\ b \end{bmatrix}}_{506 \times 1}$$

$$MSE = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}_i - y)^2$$






MSE vs Epoch with learning rate 0.0001.png
Mean Squared Loss
iteration
250
225
200
175
150
125
100
75
0
25
50
75
100
125
150
175
200